

CLAIMS

Having thus described our invention, what we claim as new and desire to secure by Letters Patent is as follows:

- 1 1. A method for resource allocation where only a marginal utility is known,
2 comprising the steps of:
3 determining an initial step size;
4 evaluating step size effectiveness with only information on the
5 marginal utility;
6 reducing the step size if necessary; and
7 deploying an effective step size for utility optimization.
- 1 2. The method of claim 1, wherein allocated resources are database memory
2 components.
- 1 3. The method of claim 1, wherein allocated resources are computer memory
2 resources.
- 1 4. The method of claim 1, wherein allocated resources are computer system
2 resources and resources are allocated to achieve optimal service level
3 objectives.
- 1 5. An apparatus for resource allocation comprising:
2 a step size determination engine;
3 an optimization module for evaluating step size effectiveness with only
4 information on the marginal utility;
5 a constraint module for reducing the step size if necessary; and

6 an interface module for deploying an effective step size for utility
7 optimization.

1 6. The apparatus of claim 5, wherein allocated resources are database memory
2 components.

1 7. The apparatus of claim 5, wherein allocated resources are computer
2 memory resources.

1 8. The apparatus of claim 5, wherein allocated resources are computer system
2 resources and resources are allocated to achieve optimal service level
3 objectives.

1 9. A method of maximizing or minimizing an objective function $f(x)$, subject
2 to constraints on a vector x where each vector x denotes a particular allocation
3 of resources and the constraints generally describe properties of the resources
4 which must be satisfied, the method maximizing or minimizing the objective
5 function $f(x)$ while satisfying the constraints on x without a knowledge of f ,
6 said method comprising the steps of:
7 starting from an initial allocation, calculating a marginal utility of said
8 allocation;
9 calculating constraint functions of said allocation;
10 applying the calculated constraint function information and marginal
11 utility information to obtain a next allocation;
12 repeating the steps of calculating a marginal utility, calculating
13 constraint functions and applying the calculated constraint function
14 information and marginal utility information until a stopping criteria is
15 satisfied; and

16 returning a locally optimal allocation of resources.

1 10. The method of maximizing or minimizing an objective function $f(x)$ as
2 recited in claim 9, wherein said marginal utility is the gradient of the function
3 f .

1 11. The method of maximizing or minimizing the objective function $f(x)$ as
2 recited in claim 9, wherein only the gradient ∇f is known and there is no
3 procedure to evaluate the objective function f .

1 12. The method of maximizing or minimizing the objective function $f(x)$ as
2 recited in claim 9, wherein only the gradient ∇f and the Hessian $\nabla^2 f$ are known
3 and there is no procedure to evaluate the objective function f .

1 13. The method of maximizing or minimizing the objective function $f(x)$ as
2 recited in claim 9, wherein the objective function is a utility function and the
3 method maximizes the utility function.

1 14. The method of maximizing or minimizing the objective function $f(x)$ as
2 recited in claim 13, wherein said utility function is time saved.

1 15. The method of maximizing or minimizing the objective function $f(x)$ as
2 recited in claim 13, wherein said utility function is utilization of computer
3 processors.

1 16. The method of maximizing or minimizing the objective function $f(x)$ as
2 recited in claim 13, wherein said utility function is a number of transactions
3 processed.

1 17. The method of maximizing or minimizing the objective function $f(x)$ as
 2 recited in claim 13, wherein said utility function is utilization of computer
 3 memory.

1 18. The method of maximizing or minimizing the objective function $f(x)$ as
 2 recited in claim 9, wherein the objective function is a cost function and the
 3 method minimizes the cost function.

1 19. The method of maximizing or minimizing the objective function $f(x)$ as
 2 recited in claim 18, wherein said cost function is power consumption.

1 20. The method of maximizing or minimizing the objective function $f(x)$ as
 2 recited in claim 18, wherein said cost function is total disk input/output time.

1 21. The method of maximizing or minimizing the objective function $f(x)$ as
 2 recited in claim 18, wherein said cost function is average system response
 3 time.

1 22. The method of maximizing or minimizing the objective function $f(x)$ as
 2 recited in claim 9, wherein a backtracking line search is implemented in which
 3 a step α is decreased until the following condition is satisfied:

$$4 \quad (\nabla f(x + \alpha p) - c_1 \nabla f(x))^T p \leq 0,$$

5 where p is a search direction, c_1 is a constant, and $(\nabla f(x + \alpha p) - c_1 \nabla f(x))^T$ is
 6 the transpose of $(\nabla f(x + \alpha p) - c_1 \nabla f(x))$.

1 23. The method of maximizing or minimizing the objective function $f(x)$ as
2 recited in claim 9, the method optimally allocating resources wherein the
3 allocated resources are database memory components.

1 24. The method of maximizing or minimizing the objective function $f(x)$ as
2 recited in claim 9, the method optimally allocating resources wherein allocated
3 resources are computer memory resources.

1 25. The method of maximizing or minimizing the objective function $f(x)$ as
2 recited in claim 9, the method optimally allocating resources wherein allocated
3 resources are computer system resources and resources are allocated to
4 achieve optimal service level objectives